

# Distributed Robust Power System State Estimation

Vassilis Kekatos, *Member, IEEE*, and Georgios B. Giannakis\*, *Fellow, IEEE*

**Abstract**—Deregulation of energy markets, penetration of renewables, advanced metering capabilities, and the urge for situational awareness, all call for system-wide power system state estimation (PSSE). Implementing a centralized estimator though is practically infeasible due to the complexity scale of an interconnection, the communication bottleneck in real-time monitoring, regional disclosure policies, and reliability issues. In this context, distributed PSSE methods are treated here under a unified and systematic framework. A novel algorithm is developed based on the alternating direction method of multipliers. It leverages existing PSSE solvers, respects privacy policies, exhibits low communication load, and its convergence to the centralized estimates is guaranteed even in the absence of local observability. Beyond the conventional least-squares based PSSE, the decentralized framework accommodates a robust state estimator. By exploiting interesting links to the compressive sampling advances, the latter jointly estimates the state and identifies corrupted measurements. The novel algorithms are numerically evaluated using the IEEE 14-, 118-bus, and a 4,200-bus benchmarks. Simulations demonstrate that the attainable accuracy can be reached within a few inter-area exchanges, while largest residual tests are outperformed.

**Index Terms**—Alternating direction method of multipliers; bad data identification; Huber's function; phasor measurement units; SCADA measurements; multi-area state estimation.

## I. INTRODUCTION

Power system state estimation (PSSE) has been traditionally performed at regional control centers with limited interaction. However, due to the deregulation of energy markets, large amounts of power are transferred over high-rate, long-distance lines spanning several control areas [11]. These so-called tie lines, originally constructed for emergency situations, are now fully operational and must be accurately monitored. Since any control area can be strongly affected by events and decisions elsewhere, independent system operators (ISOs) can no longer operate in a truly independent fashion. The ongoing penetration of renewable sources further intensifies inter-area power transfers, while their intermittent nature necessitates more frequent state acquisition. At the same time, the advances in metering infrastructure are unprecedented: phasor measurement units (PMUs) provide finely-sampled voltage and current phasors, synchronized across the grid; smart meters reach the distribution level; and networked processors are being installed throughout the grid [6], [11]. The abundance and diversity of measurements offer advanced monitoring capabilities, but processing them constitutes a major challenge, which is exacerbated in the presence of malicious data attacks and bad data [16], [1, Ch. 5-6].

There are two key issues in modernizing the power grid monitoring infrastructure: Firstly, PSSE should be performed

at the interconnection level. Yet an interconnection may include thousands of buses, while 2-3 measurements per state are typically needed. Requiring also real-time processing along with resilience to corrupted data render centralized state estimation computationally formidable. Further, a centralized approach is vulnerable and is not flexible when it comes to policy and privacy issues. Secondly, decentralizing information processing for the power grid can be performed at several hierarchies [11]: PMU measurements can be processed by phasor data concentrators (PDCs) [26]; conventional supervisory control and data acquisition (SCADA) measurements together with PDC fused data can be aggregated by the ISO; and finally, estimates from ISOs can be merged at the interconnection level. These considerations corroborate that distributed PSSE and bad data analysis are essential for realizing the smart grid vision.

Existing distributed methods for PSSE and bad data analysis are reviewed in Section II. The PSSE problem, its unique requirements and challenges are highlighted in Section III. In Section IV, a new distributed PSSE methodology is developed. Based on the alternating direction method of multipliers [2], a systematic cooperation between local control centers is enabled with unique features: it facilitates several practical PSSE formulations; it lowers the overhead for inter-area information exchanges; its convergence is guaranteed regardless of local observability or parameter tuning; and the resultant algorithm can be executed by solvers already in use at local control centers. Building on this framework, a robust decentralized estimator is derived in Section V. Different from the conventional two-step bad data analysis, the novel approach implements Huber's M-estimator [1] in a decentralized manner, while PSSE is accomplished jointly with bad data removal. Leveraging sparsity of the introduced bad data vectors, the new algorithm augments standard PSSE solvers by a few iterations. The novel robust decentralized algorithms are numerically evaluated in Section VI, and the paper is wrapped up in Section VII. Regarding notation, lower- (upper-) case boldface letters denote column vectors (matrices), and calligraphic letters stand for sets. The notation  $(\cdot)^T$  denotes transposition, while  $:=$  defines a symbol variable.

## II. RELATED WORK

Distributed solutions were pursued since the statistical formulation of PSSE [27], when it was realized that for a chain of serially interconnected areas, Kalman filter-type updates can be invoked incrementally in space [27, Part III]. For arbitrarily connected areas though, a two-level approach with a global coordinator is required [27]. Several renditions of this hierarchical approach can be found in [5], [13], [31], [11], [15]. Most of these presume local observability, meaning that

The authors are with the ECE Dept., University of Minnesota, Minneapolis, MN 55455, USA. Dr. Kekatos is also with the Computer Engrn. & Informatics Dept., University of Patras, Greece. Emails: {kekatos,georgios}@umn.edu

local states estimated excluding boundary bus measurements are uniquely identifiable. Such an assumption may not hold due to bad data removal or because PSSE is performed at a lower than the ISO level. The need for a coordinator hinders the system's reliability, while the sought algorithms may be infeasible due to computational, communication, or policy limitations.

Decentralized PSSE solutions include block Jacobi iterations [18], [4], and an approximate algorithm developed from the optimality conditions involved [10]. However, these methods assume again local observability and convergence is not always guaranteed. The auxiliary problem principle is used in [8], but several parameters must be tuned. Local observability is waived in [29], where each area is envisioned to maintain a copy of the entire high-dimensional state vector. A first-order algorithm is proposed, yet its linear convergence scales unfavorably with the interconnection size. For a review on multi-area PSSE, see also [12].

Grossly corrupted SCADA data can potentially deteriorate PSSE results. Hence, these meter readings (a.k.a. bad data) should be identified or at least detected in a measurement set. Statistical tests, such as the  $\chi^2$ - and the largest normalized residual tests, are typically employed for bad data detection and identification, respectively [25]. Both tests rely on the LS-estimated residuals and can thus run only after PSSE has been completed. Whenever a bad datum has been identified, PSSE must rerun by ignoring that datum. Alternatively, robust estimators, such as the least-absolute deviation, the least median of squares, or Huber's estimator have been considered [3], [9], [22], [1]. Towards a multi-area setup, most existing distributed PSSE methods rely on the two aforementioned tests. Even though metering reliability is improved in the smart grid realm, bad data analysis is still a major concern especially in the face of malicious data attack threats. Stealth attacks in power meter infrastructure are studied in [19], [16]. In the absence of such attacks,  $\ell_1$ -norm based methods have been developed in [16], [30], and [7].

### III. PROBLEM FORMULATION AND PRELIMINARIES

Consider an interconnected power system consisting of  $K$  control areas, where each area comprises a subset of buses supervised by its own control center. The latter is able to (i) collect the electrical measurements recorded at area buses; (ii) reliably communicate with neighboring control centers; and (iii) carry out minimal computational tasks, such as solving a (non)linear least-squares (LS) problem. As usual, measurements are assumed to be time-synchronized within and across control area footprints. A control area here is not confined to an ISO region, but it can also model entities residing at lower grid levels, such as a substation or a PDC [26]. Alternatively, a control area can be the local balancing area under a regional balancing authority's footprint; or under a micro-grid setup, a control area may degenerate even to a single bus.

Suppose that  $M_k$  measurements aggregated at the  $k$ -th area are concatenated in  $\tilde{\mathbf{z}}_k \in \mathbb{R}^{M_k}$ , and obey the model

$$\tilde{\mathbf{z}}_k = \tilde{h}_k(\mathbf{x}_k) + \tilde{\mathbf{w}}_k \quad (1)$$

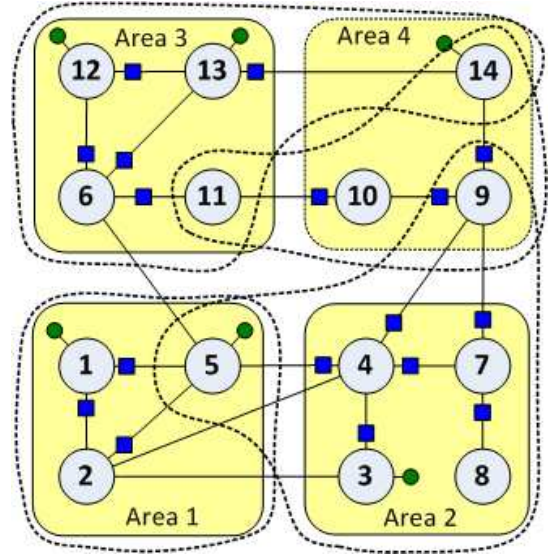


Fig. 1. The IEEE 14-bus system partitioned into four areas [28], [15]. Dotted lassos show the buses belonging to area state vectors  $\mathbf{x}_k$ . PMU bus voltage (line current) measurements depicted by green circles (blue squares).

where  $\mathbf{x}_k \in \mathbb{R}^{N_k}$  contains the subset of the interconnected power system states involved in  $\tilde{\mathbf{z}}_k$ ;  $\tilde{h}_k$  is a vector of  $M_k$  functions; and  $\tilde{\mathbf{w}}_k$  denotes a disturbance term capturing measurement error and modeling inaccuracies. Error vectors  $\{\tilde{\mathbf{w}}_k\}_{k=1}^K$  are assumed zero mean, having known covariance matrix  $\Sigma_k$ , and independent across areas. To simplify the presentation, the model of (1) can be premultiplied by  $\Sigma_k^{-1/2}$  to yield

$$\mathbf{z}_k = h_k(\mathbf{x}_k) + \mathbf{w}_k \quad (2)$$

where  $\mathbf{z}_k := \Sigma_k^{-1/2} \tilde{\mathbf{z}}_k$  and similarly for  $h(\mathbf{x}_k)$  and  $\mathbf{w}_k$ . Hence, the noise term in (2) has identity covariance matrix.

Functions  $\{h_k(\mathbf{x}_k)\}_{k=1}^K$  depend on the system's admittance matrix and are in general non-linear, except when PMUs are involved and complex quantities are expressed in rectangular representations. Performing state estimation with non-linear  $h_k$ 's entails solving non-convex optimization problems. Typically, such models are iteratively linearized via the Gauss-Newton method, or by resorting to the so called DC approximation [25], [1]. Either way, one arrives at the following computationally ubiquitous linear model (cf. (2))

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k \quad (3)$$

where  $\mathbf{H}_k \in \mathbb{R}^{M_k \times N_k}$  is known (Jacobian). When  $\{h_k\}$ s are non-linear, (3) is the linearized model per Gauss-Newton iteration.

PSSE could be performed locally at each area. Specifically, area  $k$  could aim at solving

$$\min_{\mathbf{x}_k \in \mathcal{X}_k} f_k(\mathbf{x}_k; \mathbf{z}_k, \mathbf{H}_k) \quad (4)$$

where  $f_k(\cdot)$  is a convex function of  $\mathbf{x}_k$  for the model in (3); and the convex set  $\mathcal{X}_k$  captures possible prior information, such as zero-injection buses, short circuits, or operational limits [25], [1]. Typically,  $f_k$  is chosen equal to  $\frac{1}{2} \|\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k\|_2^2$ . For this choice, the minimizer of (4) is the LS estimate (LSE),

which yields the maximum-likelihood estimate (MLE) of  $\mathbf{x}_k$  if  $\mathbf{w}_k$  is Gaussian. To derive other MLEs and/or facilitate bad data removal, alternative forms of  $f_k$  can be employed; cf. Section V. For notational simplicity, the dependence of  $f_k$  on  $\mathbf{z}_k$  and  $\mathbf{H}_k$  will be henceforth dropped.

One of the unique PSSE characteristics in interconnected areas is that the local state vectors  $\{\mathbf{x}_k\}_{k=1}^K$  overlap partially (cf. the toy interconnection of Fig. 1). Supposing that both PMU data (bus voltage and line current measurements) and interconnection states (bus voltages) are expressed in rectangular coordinates, the linear model of (3) is exact. Area 2 supervises buses  $\{3, 4, 7, 8\}$ . But since it collects current readings on lines (7, 9) and (4, 5), its state vector  $\mathbf{x}_2$  extends to the bus voltages of  $\{5, 9\}$  as well. Thus, area 2 shares the states of bus 5 (9) with area 1 (4). Notationally, let the  $N \times 1$  vector  $\mathbf{x}$  collect all the states. For every two neighboring areas, say  $k$  and  $l$ , identify the intersection of their states as  $\mathcal{S}_{kl}$ . Let also  $\mathbf{x}_k[l]$  ( $\mathbf{x}_l[k]$ ) be the sub-vector of  $\mathbf{x}_k$  ( $\mathbf{x}_l$ ) consisting of their overlapping variables ordered as they appear in  $\mathbf{x}$ . For example,  $\mathbf{x}_3[4] = \mathbf{x}_4[3]$  contain the bus voltages of  $\{11, 14\}$ .

Solving the  $K$  problems of the form (4) in isolation is clearly suboptimal, let alone the fact that control areas may be locally unobservable even if external states and their associated measurements are ignored. Disagreement on boundary bus estimates over critical tie lines is another important limitation of solving (4) on a per-area basis. On the other hand, upon defining  $\mathcal{X} := \{\mathbf{x} : \mathbf{x}_k \in \mathcal{X}_k \forall k\}$ , jointly optimizing

$$\min_{\mathbf{x} \in \mathcal{X}} \sum_{k=1}^K f_k(\mathbf{x}) \quad (5)$$

at a central control center waives all these concerns and can considerably improve estimation accuracy. Yet this comes at the expense of impractical computational and communication load, increased vulnerability, and disclosure of internal system structure. Targeting the “sweet spot” between these two extremes, a decentralized method is proposed next.

#### IV. DECENTRALIZED PSSE

Tying the local tasks of (4) into a single optimization problem equivalent to (5) can be accomplished by

$$\begin{aligned} \min_{\{\mathbf{x}_k \in \mathcal{X}_k\}} \quad & \sum_{k=1}^K f_k(\mathbf{x}_k) \\ \text{s.t.} \quad & \mathbf{x}_k[l] = \mathbf{x}_l[k], \forall l \in \mathcal{N}_k, \forall k \end{aligned} \quad (6)$$

where  $\mathcal{N}_k$  is the set of areas sharing states with area  $k$ .

The constraints of (6) force neighboring areas to consent on their shared variables, which renders problems (6) and (5) equivalent. But the same constraints couple the estimation tasks across areas. To enable a truly decentralized solution, an auxiliary variable denoted by  $\mathbf{x}_{kl} \in \mathbb{R}^{|\mathcal{S}_{kl}|}$  is introduced per pair of interacting areas  $k, l$ . To keep the notation uncluttered, symbols  $\mathbf{x}_{kl}$  and  $\mathbf{x}_{lk}$  are used interchangeably for the same variable. Then, (6) can be alternatively expressed as

$$\begin{aligned} \min_{\{\mathbf{x}_k \in \mathcal{X}_k\}, \{\mathbf{x}_{kl}\}} \quad & \sum_{k=1}^K f_k(\mathbf{x}_k) \\ \text{s.t.} \quad & \mathbf{x}_k[l] = \mathbf{x}_{kl}, \text{ for all } l \in \mathcal{N}_k, k = 1, \dots, K. \end{aligned} \quad (7)$$

The novelty here is solving (7) using the alternating direction method of multipliers (ADMM), a method that has been successfully applied for distributing several optimization problems; see e.g., [2] for a review. In ADMM, Lagrange multipliers  $\mathbf{v}_{k,l} \in \mathbb{R}^{|\mathcal{S}_{kl}|}$  are introduced for each constraint of (7). Observe that  $\mathbf{v}_{k,l}$  and  $\mathbf{v}_{l,k}$  correspond to the distinct constraints  $\mathbf{x}_k[l] = \mathbf{x}_{kl}$  and  $\mathbf{x}_l[k] = \mathbf{x}_{kl}$ , respectively. ADMM then exploits the method of multipliers concatenated with an iteration of the Gauss-Seidel algorithm. Specifically for (7), one first defines the augmented Lagrangian function

$$L(\{\mathbf{x}_k\}, \{\mathbf{x}_{kl}\}; \{\mathbf{v}_{k,l}\}) := \sum_{k=1}^K \left[ f_k(\mathbf{x}_k) + \sum_{l \in \mathcal{N}_k} \left( \mathbf{v}_{k,l}^T (\mathbf{x}_k[l] - \mathbf{x}_{kl}) + \frac{c}{2} \|\mathbf{x}_k[l] - \mathbf{x}_{kl}\|_2^2 \right) \right] \quad (8)$$

where  $c > 0$  is a predefined constant. Letting  $r$  denote the iteration index, ADMM cycles through three steps:

$$\{\mathbf{x}_k^{t+1}\} := \arg \min_{\{\mathbf{x}_k \in \mathcal{X}_k\}} L(\{\mathbf{x}_k\}, \{\mathbf{x}_{kl}^t\}; \{\mathbf{v}_{k,l}^t\}) \quad (9a)$$

$$\{\mathbf{x}_{kl}^{t+1}\} := \arg \min_{\{\mathbf{x}_{kl}\}} L(\{\mathbf{x}_k^{t+1}\}, \{\mathbf{x}_{kl}\}; \{\mathbf{v}_{k,l}^t\}) \quad (9b)$$

$$\mathbf{v}_{k,l}^{t+1} := \mathbf{v}_{k,l}^t + c (\mathbf{x}_k^{t+1}[l] - \mathbf{x}_{kl}^{t+1}) \text{ for all } k, l. \quad (9c)$$

At step (9a),  $\{\mathbf{x}_k\}$ s are updated by minimizing the augmented Lagrangian while keeping  $\mathbf{x}_{kl}$  and  $\mathbf{v}_{k,l}$  fixed to their previous iteration values;  $\mathbf{x}_{kl}$  and  $\mathbf{v}_{k,l}$  can be initialized to zero. Likewise,  $\mathbf{x}_{kl}$  are updated in (9b). Finally, (9c) is a gradient ascent of  $L(\{\mathbf{x}_k^{t+1}\}, \{\mathbf{x}_{kl}^{t+1}\}; \{\mathbf{v}_{k,l}\})$  with step size  $c$ .

Inheriting ADMM features, the minimization in (9a) decouples over control areas. Moreover, by exploiting the problem structure, the iterations (9) can be greatly simplified as presented next and detailed in the Appendix.

**Proposition 1.** *The steps in (9) yield the same  $\mathbf{x}_k^t$  iterates as the following steps*

$$\mathbf{x}_k^{t+1} := \arg \min_{\mathbf{x}_k \in \mathcal{X}_k} f_k(\mathbf{x}_k) + \frac{c}{2} \sum_{\substack{i=1 \\ \mathcal{N}_k^i \neq \emptyset}}^{N_k} |\mathcal{N}_k^i| (x_k(i) - p_k^t(i))^2, \forall k \quad (10a)$$

$$s_k^{t+1}(i) := \frac{1}{|\mathcal{N}_k^i|} \sum_{l \in \mathcal{N}_k^i} x_l^{t+1}[i], \forall i \text{ with } \mathcal{N}_k^i \neq \emptyset \quad (10b)$$

$$p_k^{t+1}(i) := p_k^t(i) + s_k^{t+1}(i) - \frac{x_k^t(i) + s_k^t(i)}{2}, \forall i, \mathcal{N}_k^i \neq \emptyset \quad (10c)$$

where  $x_k(i)$  is the  $i$ -th entry of  $\mathbf{x}_k$ ; the set  $\mathcal{N}_k^i$  consists of the areas sharing the variable  $x_k(i)$  with area  $k$ ; and  $x_l[i]$  denotes the entry of  $\mathbf{x}_l$  corresponding to  $x_k(i)$  defined for all  $l \in \mathcal{N}_k^i$ . Regarding initialization, state variables  $\mathbf{x}_k$  are set to arbitrary values  $\mathbf{x}_k^0$ ; variables  $p_k^0(i)$  are initialized to  $(x_k^0(i) + s_k^0(i))/2$ ; and  $s_k^0(i)$  as in (10b).

The minimization in (10a) and the simple update of (10c) are performed at the local centers. The averaging step of (10b) is accomplished either through a coordinator, or locally too. Either way, the information revealed per area  $k$  is *minimal*. No measurements or regression matrices, but only the boundary bus states need to be exchanged, and only between the



interested neighboring areas. States can be initialized to the flat profile [25], or some prior estimate. To implement the LSE given a non-linear measurement model [cf. (2)], the derived algorithm should be used per Gauss-Newton iteration. At each Gauss-Newton iteration, an approximate linear model is updated locally, and the decentralized iterates are initialized using the latest state estimates.

Supposing  $\{f_k(\mathbf{x}_k)\}_{k=1}^K$  are convex, and  $\{\mathcal{X}_k\}_{k=1}^K$  are closed convex sets, the cost in (7) evaluated at  $\{\mathbf{x}_k^t\}$  generated by (10) converges under mild conditions (typically met in practice) to the optimal value of (5) [2, p. 17]. Hence, when the power system is globally observable, i.e., the centralized LSE is unique, ADMM iterates minimizing the LS cost converge to it. Notwithstanding, observability is not necessary for the method to converge: if the system is unobservable, ADMM iterates converge to one of the multiple LSEs. The equivalence to the centralized LSE ensures that global observability analysis, assumed given in this work, carries over to the decentralized approach too. In addition, ADMM iterates have been shown to be resilient to asynchronous updates and random failures in the inter-area communication links [32].

For notational convenience, define per area  $k$  a diagonal matrix  $\mathbf{D}_k$  with  $(i, i)$ -th entry  $|\mathcal{N}_k^i|$ . Recall that by definition,  $|\mathcal{N}_k^i|$  is zero for strictly local states. Also, define the  $N_k$ -dimensional vector  $\mathbf{p}_k^t$  with  $i$ -th entry the  $p_k^t(i)$  of (10c) when  $|\mathcal{N}_k^i| > 0$ , and 0 otherwise. Hence, the second term in the cost of (10a) is expressed as  $\frac{c}{2} \|\mathbf{D}_k^{\frac{1}{2}}(\mathbf{x}_k - \mathbf{p}_k^t)\|_2^2$ . For the typical case of the unconstrained LSE, the minimizer of (10a) is clearly given in closed form by

$$\hat{\mathbf{x}}_k^{t+1} := (\mathbf{H}_k^T \mathbf{H}_k + c \mathbf{D}_k)^{-1} (\mathbf{H}_k^T \mathbf{z}_k + c \mathbf{D}_k \mathbf{p}_k^t) \quad (11)$$

which is a simple yet systematic modification of the local LSE. Existing PSSE software can be straightforwardly exploited for finding (11) by simply adding  $\sqrt{c} \mathbf{D}_k^{1/2} \mathbf{p}_k^t$  as pseudomeasurements with diagonal loading matrix  $\sqrt{c} \mathbf{D}_k^{1/2}$ . Note that pseudomeasurements are actually added only for the shared states. As empirically observed in Section VI, the procedure terminates after a few tens of iterations.

## V. DECENTRALIZED BAD DATA ANALYSIS

Time skews, instrument/communication failures, infrequent instrument calibration, reverse wiring, and parameter uncertainty can yield grossly corrupted SCADA measurements. In the cyber-physical smart grid context, bad data are not simply unintentional metering faults, but can also take the form of malicious data injections [24]. If an intruder can counterfeit some meters so that the attack vector lies in the range space of the PSSE regression matrix, the attack is undetectable and can arbitrarily perturb state estimates [19], [16]. Excluding these naturally termed stealth attacks, this section focuses on bad data identification. After presenting an outlier-aware estimator, interesting connections are established, to efficiently implement it using the decentralized approach of Section IV.

### A. Interconnection-Wide Bad Data Identification

The local quantities  $\{\mathbf{z}_k, \mathbf{H}_k, \mathbf{w}_k\}_{k=1}^K$  in (3) can be vertically stacked in  $\mathbf{z}$ ,  $\mathbf{H}$ , and  $\mathbf{w}$ , respectively. The

interconnection-level model then reads  $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{w}$ , where the dimensions of  $\mathbf{z}$  and  $\mathbf{x}$  are  $M = \sum_{k=1}^K M_k$  and  $N$ , in the order given. However, when bad data are present, a more pertinent model is

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{o} + \mathbf{w} \quad (12)$$

where  $\mathbf{o}$  is an unknown vector with its entry  $o(i)$  being non-zero only if  $z(i)$  is a bad datum [14], [16], [7]. Recovering both  $\mathbf{x}$  and  $\mathbf{o}$  essentially reveals the state and identifies faulty measurements. Such a mission however seems rather impossible, since the model in (12) is unobservable even if  $\mathbf{H}$  is full column rank. By capitalizing on the sparsity of  $\mathbf{o}$  though, interesting results can be obtained [14]. If  $\tau_0$  bad data are expected, then one would ideally wish to solve

$$\min_{\mathbf{x} \in \mathcal{X}, \mathbf{o}} \left\{ \frac{1}{2} \|\mathbf{z} - \mathbf{H}\mathbf{x} - \mathbf{o}\|_2^2 : \|\mathbf{o}\|_0 \leq \tau_0 \right\}. \quad (13)$$

But the  $\ell_0$ -pseudonorm, i.e., the number of non-zero  $o(i)$ 's, renders (13) NP-hard. The problem is computationally intractable even for moderate-sized interconnections and small  $\tau_0$ . Building on the premise of compressed sensing, a practical robust estimator can be derived after relaxing the  $\ell_0$ -pseudonorm by the convex  $\ell_1$ -norm as (see also [14])

$$\min_{\mathbf{x} \in \mathcal{X}, \mathbf{o}} \left\{ \frac{1}{2} \|\mathbf{z} - \mathbf{H}\mathbf{x} - \mathbf{o}\|_2^2 : \|\mathbf{o}\|_1 \leq \tau_1 \right\} \quad (14)$$

for a selected constant  $\tau_1 > 0$ , or in the Lagrangian form

$$(\hat{\mathbf{x}}, \hat{\mathbf{o}}) \in \arg \min_{\mathbf{x} \in \mathcal{X}, \mathbf{o}} \frac{1}{2} \|\mathbf{z} - \mathbf{H}\mathbf{x} - \mathbf{o}\|_2^2 + \lambda \|\mathbf{o}\|_1 \quad (15)$$

where  $\lambda$  denotes a positive parameter. The optimization problem in (15) is a convex quadratic program and can be solved by interior point-based methods. The estimator of (15) allows for joint state estimation and bad data identification. Even when some measurements are deemed as corrupted, their effect has been already suppressed, and the state estimate remains valid.

### B. Interesting Links

The two statistical tests traditionally used for bad data analysis rely on the model  $\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{w}$ , and the residual error achieved by the unconstrained LSE. The latter can be expressed as  $\mathbf{r} := \mathbf{P}\mathbf{z} = \mathbf{P}\mathbf{w}$ , where  $\mathbf{P} := \mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$  is the so called ‘‘residual sensitivity matrix’’ satisfying  $\mathbf{P} = \mathbf{P}^2$  [1, Ch. 5]. Clearly, when  $\mathbf{w}$  is Gaussian,  $\mathbf{r}$  is Gaussian too with covariance matrix  $\mathbf{P}$ . The  $\chi^2$ -test compares  $\|\mathbf{r}\|_2^2$  against a threshold to detect the presence of bad data [25], [1]. The largest normalized residual (LNR) test computes

$$\bar{r}_{\max} := \max_{i \in \{1, \dots, M\}} \frac{|r(i)|}{\sqrt{P(i, i)}} \quad (16)$$

where  $P(i, i)$  is the  $(i, i)$ -th entry of  $\mathbf{P}$ ; note that in the absence of critical measurements,  $0 < P(i, i) \leq 1$  for all  $i$ . Metric  $\bar{r}_{\max}$  is then compared to a prescribed threshold to identify a single bad datum [1, Sec. 5.7]. Adopting the proof in [16, Prop. 1], the following claim can be established.

**Proposition 2.** *The optimization problem in (13) with  $\tau_0 = 1$  over  $\mathcal{X} = \mathbb{R}^N$ , and the LNR test of (16) are equivalent in terms of the measurement being detected as bad.*

For multiple bad data, this connection is unclear. If a measurement is deemed as outlier, PSSE is repeated after discarding this bad datum, the LNR test is re-applied, and the process iterates till no corrupted data are identified. Even though rank-one updates can be used to speed up computations, the process becomes complicated for multi-area grids.

Returning to the convex relaxation (15) for  $\mathcal{X} = \mathbb{R}^N$ , note that when  $\lambda \rightarrow \infty$ , the minimizer  $\hat{\mathbf{o}}$  becomes zero, and thus,  $\hat{\mathbf{x}}$  reduces to the LSE. On the contrary, by letting  $\lambda \rightarrow 0^+$  [17], the solution  $\hat{\mathbf{x}}$  coincides with  $\arg \min_{\mathbf{x}} \|\mathbf{z} - \mathbf{H}\mathbf{x}\|_1$ , meaning the least-absolute value estimator [3], [9].

For finite  $\lambda > 0$ ,  $\hat{\mathbf{x}}$  of (15) corresponds to Huber's M-estimator; see [14] and references therein. Based on this connection and assuming Gaussian  $\mathbf{w}$ , tuning parameter  $\lambda$  can be set to 1.34, which makes the estimator 95% asymptotically efficient for bad data-free measurements [21, p. 26].

Alternatively, Huber's estimate can be expressed as the  $\mathbf{x}$ -minimizer of  $\min_{\mathbf{x} \in \mathcal{X}, \omega} \frac{1}{2} \|\omega\|_2^2 + \lambda \|\mathbf{z} - \mathbf{H}\mathbf{x} - \omega\|_1$  as shown in [20]. The bad data identification performance of this minimization is analyzed in [30].

To solve (15), one can first minimize over  $\mathbf{x}$  and then over  $\mathbf{o}$ . For  $\mathcal{X} = \mathbb{R}^N$ , the  $\mathbf{x}$  minimizing (15) is  $(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\mathbf{z} - \mathbf{o})$ , and thus, minimizing (15) reduces to [16], [7]

$$\min_{\mathbf{o}} \frac{1}{2} \|\mathbf{P}(\mathbf{z} - \mathbf{o})\|_2^2 + \lambda \|\mathbf{o}\|_1. \quad (17)$$

A minimization similar to (17) is derived in [16] using a generalized likelihood ratio test. By assuming a Bayesian prior  $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \Sigma_x)$ , [16] suggests solving (17), but with  $\mathbf{P}$  substituted by  $\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H} + \Sigma_x^{-1})^{-1} \mathbf{H}^T$ . In any case, matrix  $\mathbf{P}$  couples the minimization over  $\mathbf{o}$  across areas and complicates a decentralized implementation. An efficient distributed algorithm for solving (15) is developed next.

### C. Distributed Robust Algorithm

Consider now the system-wide minimization in (15) under the decentralized PSSE format of Section IV. To this end, partition  $\mathbf{o}$  into subvectors  $\mathbf{o}_k$ 's conforming to the partition of  $\mathbf{z}$  into  $\mathbf{z}_k$ 's, and define the local functions

$$f_k(\mathbf{x}_k, \mathbf{o}_k) := \frac{1}{2} \|\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k - \mathbf{o}_k\|_2^2 + \lambda \|\mathbf{o}_k\|_1. \quad (18)$$

Critically, notice that  $\{\mathbf{o}_k\}$  belong to a single area and there is no need for sharing them. Similar to the way (7) was obtained from (5), the decentralized equivalent of (15) is

$$\begin{aligned} \min_{\{\mathbf{x}_k \in \mathcal{X}_k\}, \{\mathbf{o}_k\}, \{\mathbf{x}_{kl}\}} & \sum_{k=1}^K f_k(\mathbf{x}_k, \mathbf{o}_k) \\ \text{s.t. } & \mathbf{x}_k[l] = \mathbf{x}_{kl}, \text{ for all } l \in \mathcal{N}_k, k = 1, \dots, K. \end{aligned} \quad (19)$$

Proceeding with the ADMM methodology, the augmented Lagrangian of (19) is similar to the one in (8) apart from  $f_k(\mathbf{x}_k)$  being replaced by  $f_k(\mathbf{x}_k, \mathbf{o}_k)$ . Having three instead of two primal variable sets offers two algorithmic alternatives: the additional  $\mathbf{o}_k$  variables can be jointly optimized either with  $\{\mathbf{x}_k\}$ s in step (9a), or with  $\{\mathbf{x}_{kl}\}$ s in step (9b). The second choice yields computational efficiency as described next.

Two are the key observations here. First, that the augmented Lagrangian of the problem is separable with respect to  $\{\mathbf{o}_k\}$

---

### Algorithm 1 Decentralized Robust PSSE (D-RPSSE)

---

**Require:**  $\{\mathbf{H}_k, \mathbf{D}_k, \mathbf{z}_k\}$ , and positive  $c, \lambda$ .

- 1: Initialize  $\{\mathbf{x}_k^0, \mathbf{p}_k^0, \mathbf{s}_k^0\}$  as in (10); and  $\{\mathbf{o}_k^0\}$  to zero.
  - 2: **for**  $t = 1, 2, \dots$  **do**
  - 3:   Every area updates its  $\mathbf{x}_k^{t+1}$  as in (20).
  - 4:   Neighboring areas exchange shared state variables.
  - 5:   Every area updates its  $\mathbf{s}_k^{t+1}$  via (10b).
  - 6:   Every area updates its  $\mathbf{p}_k^{t+1}$  via (10c).
  - 7:   Every area updates its  $\mathbf{o}_k^{t+1}$  through (21)-(22).
  - 8: **end for**
- 

and  $\{\mathbf{x}_{kl}\}$ , and hence, the optimization required at the ADMM step (9b) decouples over the two variable sets. Second, the updated  $\{\mathbf{o}_k^{t+1}\}$  vectors appear only in  $f_k(\mathbf{x}_k, \mathbf{o}_k^t)$ , and do not affect the  $\mathbf{x}_{kl}^{t+1}$  and  $\mathbf{v}_{k,l}^{t+1}$  updates. Since the iterations of (10) were derived from (9) for generic  $\{f_k(\mathbf{x}_k)\}$ , they can be readily extended to the robust PSSE case as follows.

In step (10a),  $\{f_k(\mathbf{x}_k)\}$  should be replaced by  $\{f_k(\mathbf{x}_k, \mathbf{o}_k^t)\}$ . For  $\mathcal{X}_k = \mathbb{R}^{N_k}$  and using the notation introduced before (11), state variables can be updated in closed form as

$$\mathbf{x}_k^{t+1} := (\mathbf{H}_k^T \mathbf{H}_k + c \mathbf{D}_k)^{-1} (\mathbf{H}_k^T (\mathbf{z}_k - \mathbf{o}_k^t) + c \mathbf{D}_k \mathbf{p}_k^t). \quad (20)$$

Interestingly, the update of (20) is the LSE of (11) slightly modified: measurements  $\mathbf{z}_k$  have been substituted by the bad data-compensated measurements  $(\mathbf{z}_k - \mathbf{o}_k^t)$ .

Steps (10b)-(10c) are left untouched; guaranteeing the consent between shared states is unrelated to the local functions  $f_k(\mathbf{x}_k)$ . The variables  $\{\mathbf{o}_k\}$  can be finally updated as the minimizers of  $f_k(\mathbf{x}_k^{t+1}, \mathbf{o}_k)$  as required by the ADMM step (9b). Interestingly, the solution of the latter minimization is provided in closed form too as (cf. [2, p. 32])

$$\mathbf{o}_k^{t+1} := [\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k^{t+1}]_{\lambda}^+ \quad (21)$$

where  $[x]_{\lambda}^+$  denotes the simple thresholding operator

$$[x]_{\lambda}^+ := \begin{cases} x + \lambda, & x < -\lambda \\ 0, & |x| \leq \lambda \\ x - \lambda, & x > \lambda \end{cases} \quad (22)$$

understood entry-wise in (21). Intuitively, for measurements with a small tentative absolute residual, the corresponding  $\mathbf{o}_k(i)$  becomes zero. Otherwise, the bad datum residual is artificially shrunk by  $\lambda$  towards zero via a non-zero  $\mathbf{o}_k(i)$ .

The novel robust decentralized algorithm, called D-RPSSE hereafter, is tabulated as Alg. 1. Compared to the decentralized LSE of (10)-(11), D-RPSSE maintains software compatibility too. On top of adding  $\sqrt{c} \mathbf{D}_k^{1/2} \mathbf{p}_k^t$  as pseudo-measurements, resilience to bad data comes by simply off-setting local measurements by  $\mathbf{o}_k^t$  and via the thresholding rule of (22). Robust state estimates and bad data identification are jointly acquired without repeating PSSE as required by the LNRT.

## VI. SIMULATED TESTS

The developed decentralized state estimators are numerically tested on an Intel Duo Core @ 2.2 GHz (4GB RAM) computer using MATLAB. Two power network benchmarks

TABLE I  
AVERAGE STANDARD DEVIATION PER STATE

Estimator	IEEE 14-bus grid	IEEE 118-bus grid
Internal LSE	$3.4 \cdot 10^{-3}$	$4.1 \cdot 10^{-4}$
Local LSE	$3.1 \cdot 10^{-3}$	$4.0 \cdot 10^{-4}$
Global LSE	$1.0 \cdot 10^{-3}$	$2.2 \cdot 10^{-4}$

are initially considered, namely the IEEE 14- and the 118-bus systems; while later a 4,200-bus is generated based on the IEEE 14- and 300-bus systems [28]. Their admittance matrices and the underlying power system states are obtained using MATPOWER [33]. For all systems, the state vector contains the real and imaginary parts of all bus voltages. Measurements consist of PMU recordings on bus voltages and line currents, expressed in rectangular coordinates too. Measurement noise is simulated as independent zero-mean Gaussian with standard deviation per real component 0.01 and 0.02 for voltages and currents, respectively [33].

For the IEEE 14-bus grid, PMU sites and types are shown in Fig. 1: 6 bus voltage and 17 line current meters yield a total of 46 measurements corresponding to a redundancy ratio of 1.6 [25]. For the IEEE 118-bus grid, PMU sites are selected uniformly at random: 77 bus voltage and 205 line current meters are utilized, yielding a redundancy ratio of 2.4. The IEEE 14-bus grid is partitioned into the 4 areas depicted in Fig. 1, while the IEEE 118-bus system is split into 3 areas as in [23, Fig. 4].

A reasonable question is whether interconnection-wide PSSE offers any improvement over local PSSE. To this end, three estimators are numerically compared: First, an estimator that uses only the measurements related to its own area, henceforth called “internal.” Second, a “local” estimator which extends its state to boundary buses that can be reached via tie line measurements. Lastly, the interconnection-wide or “global” estimator. The average standard deviation per state is empirically computed over 100 Monte Carlo runs. Table I lists the standard deviations for the two power grids. The IEEE 118-bus grid attains improved estimation accuracy due to its increased redundancy ratio. More importantly, the improvement of the local over the internal estimator is marginal, whereas the accuracy of the global estimator roughly doubles. This observation speaks for the importance of interconnection-wide PSSE even when local observability is guaranteed.

#### A. Testing the Decentralized LSE

States are initialized here to the flat profile. Even though iterations (10) are guaranteed to converge to the optimal value of (5) for any  $c > 0$ , the value of  $c$  affects the convergence rate. After scaling the data to obey the model in (3),  $c$  is empirically set to  $10^4$ . Two performance metrics are adopted: the per area error to the centralized solution of (5), denoted by  $e_{k,c}^t := \|\mathbf{x}_k^{(c)} - \mathbf{x}_k^t\|_2 / N_k$ , and the per area error to the true underlying state defined as  $e_{k,o}^t := \|\mathbf{x}_k - \mathbf{x}_k^t\|_2 / N_k$ .

Fig. 2 depicts the  $e_{k,c}^t$  and the  $e_{k,o}^t$  curves obtained for the IEEE 14-bus network. The almost flat  $e_{k,c}^t$  curves shown at the top of the figure correspond to the decentralized algorithm of [29] whose step sizes have been optimized. Based on the

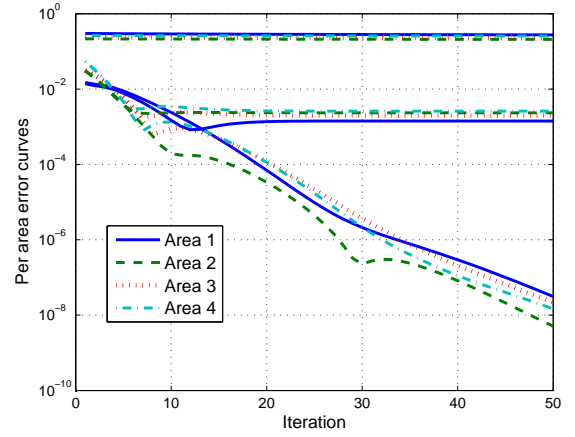


Fig. 2. Per area error curves  $e_{k,c}^t$ 's (bottom) and  $e_{k,o}^t$ 's (middle) for the decentralized LSE of the IEEE 14-bus system of Fig. 1. The almost flat curves (top) show the corresponding  $e_{k,c}^t$  error curves for the algorithm of [29].

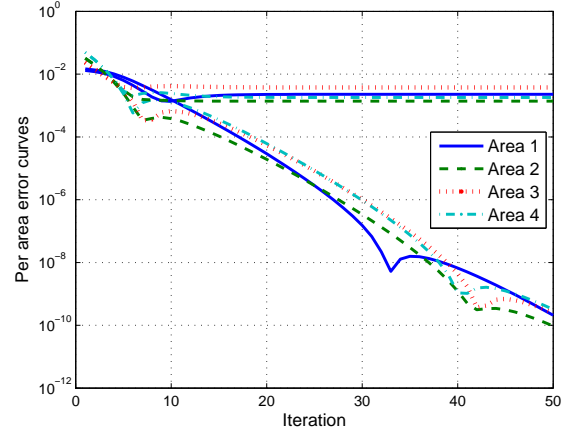


Fig. 3. Per area error curves  $e_{k,c}^t$ 's (bottom) and  $e_{k,o}^t$ 's (top) for the decentralized LSE of the IEEE 14-bus system without local observability.

$e_{k,c}^t$  curves, the algorithm of (10) converges to the centralized solution. Interestingly though, as indicated by the  $e_{k,o}^t$  curves, accuracy of around  $10^{-3}$  dictated by the measurements is reached in 10-15 iterations. On the other hand, the algorithm of [29] attains the same accuracy after around 10,000 iterations. Being a first-order method, the algorithm in [29] incurs per iteration complexity much smaller than (10), yet it does not fully exploit the capabilities of local PSSE solvers. Moreover, the high number of iterations required translates to increased inter-area communication overhead.

To evaluate the new algorithm in scenarios where local observability does not hold, the electric current measurement on line (6, 11) is removed from the IEEE 14-bus measurement set (cf. Fig. 1). Since the only measurement directly related to bus 11 is the current measurement on line (10, 11) and that is collected by control area 4, area 3 is locally unobservable. The error curves obtained and plotted in Fig. 3 verify that the developed method does not require local observability.

Switching to the IEEE 118-bus benchmark, similar results are observed. As evidenced by the  $e_{k,c}^t$  and  $e_{k,o}^t$  curves plotted in Fig. 4, the decentralized solution attains the desired statistical accuracy within only 5-10 iterations.

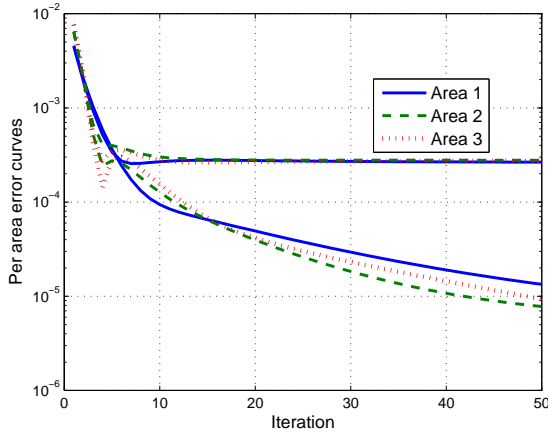


Fig. 4. Per area error curves  $e_{k,c}^t$ 's (bottom) and  $e_{k,o}^t$ 's (top) for the decentralized LSE of the IEEE 118-bus system of [23, Fig. 4].

TABLE II  
MEAN-SQUARE ESTIMATION ERROR IN THE PRESENCE OF BAD DATA

Method	GA-LSE	LSE	LNRT	Huber's
(S0)	0.0278	0.0278	0.0286	0.0281
(S1)	0.0313	0.0318	0.0331	0.0322
(S2)	0.0336	0.1431	0.0404	0.0390
(S3)	0.0367	0.1434	0.0407	0.0390

### B. Testing the Decentralized Robust Estimator

The centralized versions of bad data analysis methods are compared first. The IEEE 14-bus grid of Fig. 1 is considered under the following four scenarios. **(S0)**: no bad data; **(S1)**: bad data on line (4, 7); **(S2)**: bad data on line current (4, 7) and bus voltage 5; and **(S3)**: bad data on line current (4, 7), tie line current (10, 11), and bus voltage 5. In all scenarios, bad data are simulated by multiplying the real and imaginary parts of the actual measurement by 1.2. The performance metric here is the  $\ell_2$ -norm between the true state and the PSSE, which is averaged over 1,000 Monte Carlo runs.

Table II lists the results obtained by the four centralized algorithms tested: (a) an ideal yet practically infeasible genie-aided LSE (GS-LSE), which ignores the corrupted measurements; (b) the regular LSE; (c) the LNR test-based (LNRT) estimator with the test threshold set to 3.0 [1]; and (d) Huber's estimator of (15) with  $\lambda = 1.34$  and  $\mathcal{X} = \mathbb{R}^N$ . For (S0)-(S1), the estimators perform almost similarly. The few corrupted measurements in (S2)-(S3) can deteriorate LSE's performance, while Huber's estimator performs slightly better than LNRT. Computationally, Huber's estimator is implemented using iterations (20)-(21) for the interconnection-wide vectors  $\mathbf{x}$  and  $\mathbf{o}$  with  $c = 0$ . The algorithm is terminated when the  $\ell_2$ -norm between the two last state iterates becomes less than  $10^{-4}$ . On the average and for all scenarios (S0)-(S3), Huber's estimator converges in 6-12 iterations and within 1.3 msec, while LNRT requires 2.6 recalculations of (16) in 1.5 msec. The computing times were also measured for the IEEE 118-bus grid without corrupted data. Interestingly, the average time on the IEEE 118-bus grid without corrupted data are 3.2 msec and 81 msec, respectively. Of course, efficient updates for LNRT can be devised, but their decentralized implementation is not obvious.

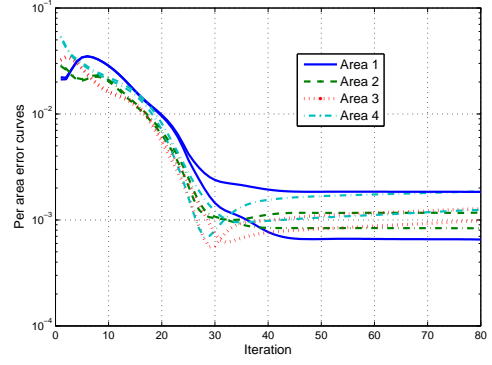


Fig. 5. Per area error curves  $e_{k,c}^t$ 's (bottom) and  $e_{k,o}^t$ 's (top) for the D-RPSSE algorithm on the IEEE 14-bus benchmark under (S3).

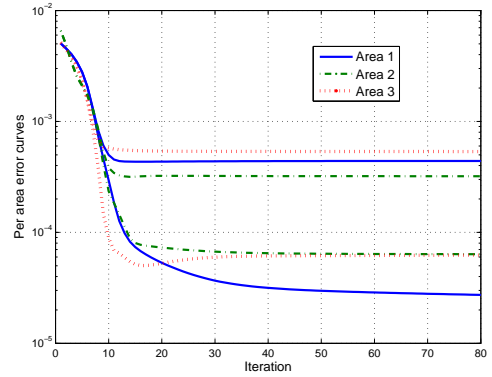


Fig. 6. Per area error curves  $e_{k,c}^t$ 's (bottom) and  $e_{k,o}^t$ 's (top) for the D-RPSSE algorithm on the IEEE 118-bus benchmark having 10% of the measurements corrupted.

Focusing on decentralized implementation, the D-RPSSE algorithm (cf. Alg. 1) is considered next. Technically, the  $\hat{\mathbf{x}}$  minimizing (15) may not be unique. However, a sufficient condition for its uniqueness is provided in [17, Th. 3.6], and this condition was satisfied in all the remaining tests. D-RPSSE was first tested on the IEEE 14-bus grid under scenario (S3). The associated  $e_{k,c}^t$  and  $e_{k,o}^t$  curves are depicted in Fig. 5. The decentralized iterates approach the underlying state at an accuracy of  $10^{-3}$  in 30 iterations in 12.1 msec. In comparison, the respective time for the LSE was 5 msec. Finally, for the IEEE 118-bus system, 10% of the measurements are corrupted in the way described earlier. The corresponding error curves are plotted in Fig. 6. The time needed to achieve a  $10^{-3}$ – $10^{-4}$  accuracy (10 iterations) is 12.1 msec versus 3.7 msec for the related LSE.

The decentralized algorithms were finally tested on a larger power network: a 4,200-bus power grid that was generated using the IEEE 14- and 300-bus grids as follows. Each one of the 300 buses of the latter is assumed to be a different area, and is replaced by a copy of the IEEE 14-bus grid. Additionally, every branch of the IEEE 300-bus grid is now an inter-area line whose terminal buses are randomly selected from the two incident to this line areas. Measurements, bad data, and  $c$  are selected as in the tests for the IEEE 118-bus grid.



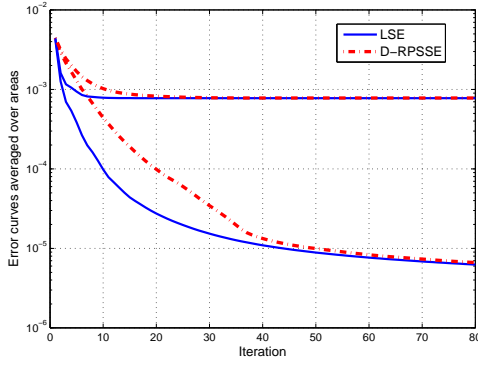


Fig. 7. Average error curves  $\sum_{k=1}^{300} e_{k,c}^t/300$  (bottom) and  $\sum_{k=1}^{300} e_{k,o}^t/300$  (top) for the LSE and D-RPSSE algorithms on a 4,200-bus grid.

Fig. 7 shows the corresponding error curves that are now averaged over the 300 areas. Given bad data-free measurements, the decentralized LSE approaches the underlying state at an accuracy of  $10^{-3}$  in approximately 10 iterations or 6.2 msec; the centralized LSE finished in 93.4 msec. For a 10% of the measurements being bad, D-RPSSE yields an accuracy of  $10^{-3}$  in less than 20 ADMM iterations or 15.2 msec; the centralized robust estimator needed 193.5 msec. The convergence to their unique centralized counterparts is illustrated by the bottom curves. These tests corroborate that (i) the decentralized algorithms are basically insensitive to the variation of  $c$ ; and (ii) the convergence time for both the decentralized LSE and D-RPSSE scale favorably with the network size. It is worth mentioning that the reverse topology where each of the 14 nodes of the IEEE 14-bus grid is replaced by a IEEE 300-bus system was tested too. Under this second architecture, the algorithms converged even faster due to the looser area coupling.

## VII. CONCLUSIONS

Distributed and robust state estimators have been treated here under a systematic manner. The proposed algorithms waive local observability requirements and maintain backward compatibility. With a few minimal data exchanges between neighboring areas, local control centers can acquire highly accurate estimates for the part of the interconnection they are responsible for, and simultaneously identify (un)intentionally corrupted data. The novel framework accommodates several important modifications of the PSSE problem, such as constraints (e.g., zero-injection buses, operational limits) and different MLEs. Exciting issues that emerge for future research include this work's application to generalized state estimation, its applicability to non-convex problems, and re-weighted versions of (15) [14].

## APPENDIX

A useful lemma is shown first.

**Lemma 1.** *For every pair of adjacent areas  $k$  and  $l$ , the Lagrange multipliers updated by (9c) satisfy  $\mathbf{v}_{k,l}^t + \mathbf{v}_{l,k}^t = \mathbf{0}$  per iteration  $r > 0$ .*

*Proof:* Note that step (9b) decouples over the  $\mathbf{x}_{kl}$ 's, while the minimizers can be shown to be

$$\mathbf{x}_{kl}^{t+1} := \left( \frac{\mathbf{x}_k^{t+1}[l] + \mathbf{x}_l^{t+1}[k]}{2} \right) + \left( \frac{\mathbf{v}_{k,l}^t + \mathbf{v}_{l,k}^t}{2c} \right). \quad (23)$$

Next, consider the updates of  $\mathbf{v}_{k,l}$  and  $\mathbf{v}_{l,k}$  according to step (9c). Adding the two updates by parts and solving for the common term  $\mathbf{x}_{kl}^{t+1}$ , yields

$$\mathbf{x}_{kl}^{t+1} = \frac{\mathbf{v}_{k,l}^t + \mathbf{v}_{l,k}^t}{2c} - \frac{\mathbf{v}_{k,l}^{t+1} + \mathbf{v}_{l,k}^{t+1}}{2c} + \frac{\mathbf{x}_k^{t+1}[l] + \mathbf{x}_l^{t+1}[k]}{2}.$$

By equating the right-hand sides of (23) and the last equation, the claim of the lemma follows readily. ■

*Proof of Proposition 1:* The optimization in (9a) is separable across areas. Upon completing the squares, the optimization task for area  $k$  during step (9a) becomes

$$\min_{\mathbf{x}_k} f_k(\mathbf{x}_k) + \frac{c}{2} \sum_{l \in \mathcal{N}_k} \left\| \mathbf{x}_k[l] - \left( \mathbf{x}_{kl}^t - \frac{\mathbf{v}_{k,l}^t}{c} \right) \right\|_2^2. \quad (24)$$

Apparently, the  $\ell_2$ -norms in (24) decouple over the entries of the vectors involved. However, a single entry of  $\mathbf{x}_k$ , say  $x_k(i)$ , may be shared not only between areas  $k$  and  $l$ , but rather among area  $k$  and all the areas in  $\mathcal{N}_k^i$ . If  $x_{kl}[i]$  ( $v_{k,l}[i]$ ) denotes the entry of  $\mathbf{x}_{kl}$  ( $\mathbf{v}_{k,l}$ ) corresponding to  $x_k(i)$ , the optimization in (24) can be expressed as

$$\min_{\mathbf{x}_k} f_k(\mathbf{x}_k) + \frac{c}{2} \sum_{\substack{i \in \mathcal{N}_k \\ \mathcal{N}_k^i \neq \emptyset}} |\mathcal{N}_k^i| (x_k(i) - p_k^{t+1}(i))^2 \quad (25)$$

where for all  $k$ , and  $i = 1, \dots, N_k$  with  $\mathcal{N}_k^i \neq \emptyset$ , define

$$p_k^{t+1}(i) := \frac{1}{|\mathcal{N}_k^i|} \sum_{l \in \mathcal{N}_k^i} \left( x_{kl}^t[i] - \frac{v_{k,l}^t[i]}{c} \right). \quad (26)$$

By Lemma 1, step (9b) simplifies to  $\mathbf{x}_{kl}^{t+1} = 0.5 (\mathbf{x}_k^{t+1}[l] + \mathbf{x}_l^{t+1}[k])$ . In other words, the auxiliary variable  $\mathbf{x}_{kl}$  is the average of the shared state variables across areas  $k$  and  $l$  per iteration. By eliminating the auxiliary variables  $\mathbf{x}_{kl}$  from the updates of (26) and (9c), step (9b) can be dropped. Hence, one arrives at the iterates

$$\mathbf{x}_k^{t+1} := \arg \min_{\mathbf{x}_k} f_k(\mathbf{x}_k) + \frac{c}{2} \sum_{\substack{i \in \mathcal{N}_k \\ \mathcal{N}_k^i \neq \emptyset}} |\mathcal{N}_k^i| (x_k(i) - p_k^t(i))^2 \quad (27a)$$

$$\mathbf{v}_{k,l}^{t+1} := \mathbf{v}_{k,l}^t + c \left( \frac{\mathbf{x}_k^{t+1}[l] - \mathbf{x}_l^{t+1}[k]}{2} \right) \quad (27b)$$

$$p_k^{t+1}(i) := \frac{1}{2} \left( x_k^{t+1}(i) + \frac{1}{|\mathcal{N}_k^i|} \sum_{l \in \mathcal{N}_k^i} x_l^{t+1}[i] \right) - \frac{1}{|\mathcal{N}_k^i|} \sum_{l \in \mathcal{N}_k^i} \frac{v_{k,l}^{t+1}[i]}{c}. \quad (27c)$$

To further simplify the iterations, define the average of the shared variable  $x_k(i)$ 's copies over  $\mathcal{N}_k^i$  as the  $s_k^{t+1}(i)$  in (10b). Define also the average of the weighted Lagrange multipliers  $u_k^{t+1}(i) := \sum_{l \in \mathcal{N}_k^i} v_{k,l}^{t+1}[i] / (c|\mathcal{N}_k^i|)$ . Then, (27c) can be written as

$$p_k^{t+1}(i) := \frac{1}{2} (x_k^{t+1}(i) + s_k^{t+1}(i)) - u_k^{t+1}(i). \quad (28)$$



With  $\{v_{k,l}\}$  initialized to zero,  $\{u_k(i)\}$  can be recursively updated as  $u_k^{t+1}(i) := u_k^t(i) + (x_k^{t+1}(i) - s_k^{t+1}(i))/2$ . Hence, update (28) can be alternatively performed as in (10c). Collecting (27a), the definition of  $\{s_k^t(i)\}$ , and the recursive updates for  $\{p_k^t(i)\}$ , one readily arrives at (10). ■

## REFERENCES

- [1] A. Abur and A. Gomez-Exposito, *Power System State Estimation: Theory and Implementation*. New York, NY: Marcel Dekker, 2004.
- [2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learning*, vol. 3, pp. 1–122, 2010.
- [3] M. K. Celik and A. Abur, "A robust WLAV state estimator using transformations," *IEEE Trans. Power Syst.*, vol. 7, no. 1, pp. 106–113, Feb. 1992.
- [4] A. J. Conejo, S. de la Torre, and M. Canas, "An optimization approach to multiarea state estimation," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 213–221, Feb. 2007.
- [5] T. V. Cutsem and M. Ribbens-Pavella, "Critical survey of hierarchical methods for state estimation of electric power systems," *IEEE Trans. Power App. Syst.*, vol. 102, no. 10, pp. 3415–3424, Oct. 1983.
- [6] J. De La Ree, V. Centeno, J. Thorp, and A. Phadke, "Synchronized phasor measurement applications in power systems," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 20–27, Jun. 2010.
- [7] D. Duan, L. Yang, and L. L. Scharf, "Phasor state estimation from PMU measurements with bad data," in *Proc. IEEE Workshop on Comp. Adv. in Multi-Sensor Adaptive Proc.*, San Juan, Puerto Rico, Dec. 2011.
- [8] R. Ebrahimian and R. Baldick, "State estimation distributed processing," *IEEE Trans. Power Syst.*, vol. 15, no. 4, pp. 1240–1246, Nov. 2000.
- [9] A. A. El-Keib, J. Nieplocha, H. Singh, and D. Maratukulam, "A decomposed state estimation technique suitable for parallel processor implementation," *IEEE Trans. Power Syst.*, vol. 7, no. 3, pp. 1088–1097, Aug. 1992.
- [10] D. M. Falcao, F. F. Wu, and L. Murphy, "Parallel and distributed state estimation," *IEEE Trans. Power Syst.*, vol. 10, no. 2, pp. 724–730, May 1995.
- [11] A. Gómez-Exposito, A. Abur, A. de la Villa Jaén, and C. Gómez-Quiles, "A multilevel state estimation paradigm for smart grids," *Proc. IEEE*, vol. 99, no. 6, pp. 952–976, Jun. 2011.
- [12] A. Gómez-Exposito, A. de la Villa Jaén, C. Gómez-Quiles, P. Rousseaux, and T. V. Cutsem, "A taxonomy of multi-area state estimation methods," *Electric Power Systems Research*, vol. 81, pp. 1060–1069, 2011.
- [13] S. Iwamoto, M. Kusano, and V. H. Quintana, "Hierarchical state estimation using a fast rectangular-coordinate method," *IEEE Trans. Power Syst.*, vol. 4, no. 3, pp. 870–880, Aug. 1989.
- [14] V. Kekatos and G. B. Giannakis, "From sparse signals to sparse residuals for robust sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3355–3368, Jul. 2011.
- [15] G. N. Korres, "A distributed multiarea state estimation," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 73–84, Feb. 2011.
- [16] O. Kosut, L. Jia, J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 645–658, Dec. 2011.
- [17] W. Li and J. J. Swetits, "The linear  $\ell_1$  Estimator and Huber M-Estimator," *SIAM J. Optim.*, vol. 8, pp. 457–475, May 1998.
- [18] S.-Y. Lin and C.-H. Lin, "An implementable distributed state estimator and distributed bad data processing schemes for electric power systems," *IEEE Trans. Power Syst.*, vol. 9, no. 3, pp. 1277–1284, Aug. 1994.
- [19] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *Proc. ACM Conf. on Computer and Comm. Security*, Chicago, IL, Nov. 2009, pp. 21–32.
- [20] O. L. Mangasarian and D. R. Musicant, "Robust linear and support vector regression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 9, pp. 950–955, Sep. 2000.
- [21] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*. Wiley, 2006.
- [22] L. Mili, M. G. Cheniae, and P. J. Rousseeuw, "Robust state estimation of electric power systems," *IEEE Trans. Circuits Syst. I*, vol. 41, no. 5, pp. 349–358, May 1994.
- [23] L. Min and A. Abur, "Total transfer capability computation for multi-area power systems," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1141–1147, Aug. 2006.
- [24] B. Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee *et al.*, "Cyberphysical security of a smart grid infrastructure," *Proc. IEEE*, vol. 100, no. 1, pp. 195–209, Jan. 2012.
- [25] A. Monticelli, "Electric power system state estimation," *Proc. IEEE*, vol. 88, no. 2, pp. 262–282, Feb. 2000.
- [26] A. G. Phadke and J. S. Thorp, *Synchronized Phasor Measurements and Their Applications*. New York, NY: Springer, 2008.
- [27] F. C. Schweppe, J. Wildes, and D. Rom, "Power system static state estimation: Parts I, II, and III," *IEEE Trans. Power App. Syst.*, vol. 89, pp. 120–135, Jan. 1970.
- [28] Power systems test case archive. University of Washington. [Online]. Available: <http://www.ee.washington.edu/research/pstca/>
- [29] L. Xie, D.-H. Choi, and S. Kar, "Cooperative distributed state estimation: Local observability relaxed," in *Proc. IEEE PES General Meeting*, Detroit, MI, Jul. 2011.
- [30] W. Xu, M. Wang, and A. Tang, "Sparse recovery from nonlinear measurements with applications in bad data detection for power networks," submitted, Dec. 2011. [Online]. Available: <http://arxiv.org/abs/1112.6234>
- [31] L. Zhao and A. Abur, "Multiarea state estimation using synchronized phasor measurements," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 611–617, May 2005.
- [32] H. Zhu, G. B. Giannakis, and A. Cano, "Distributed in-network decoding," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3970–3983, Oct. 2009.
- [33] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER: steady-state operations, planning and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.



and wireless communications.



His general interests span the areas of communications, networking and statistical signal processing - subjects on which he has published more than 325 journal papers, 525 conference papers, 20 book chapters, two edited books and two research monographs. Current research focuses on compressive sensing, cognitive radios, network coding, cross-layer designs, wireless sensors, social and power grid networks. He is the (co-)inventor of 21 patents issued, and the (co-)recipient of eight paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize Paper Award in Wireless Communications. He also received Technical Achievement Awards from the SP Society (2000), from EURASIP (2005), a Young Faculty Teaching Award, and the G. W. Taylor Award for Distinguished Research from the University of Minnesota. He is a Fellow of EURASIP, and has served the IEEE in a number of posts, including that of a Distinguished Lecturer for the IEEE-SP Society.

**Vassilis Kekatos** (M'10) received the Diploma, MSc., and PhD in Computer Engr. & Informatics Dept. from the University of Patras, Greece, in 2001, 2003, and 2007, respectively. Since 2009, he has been a Marie Curie Fellow, and he is currently a post doctoral associate with the Dept. of Electrical and Computer Engr. at the University of Minnesota, and the Computer Engr. & Informatics Dept. at the University of Patras, Greece. His research interests lie in the areas of statistical signal processing with emphasis on the power grid, compressive sampling,